PATENT
450100-02919

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION FOR LETTERS PATENT

TITLE: SYNCHRONIZATION CONTROL APPARATUS AND METHOD, AND RECORDING MEDIUM

INVENTORS: Keiichi YAMADA, Kenichiro KOBAYASHI, Tomoaki NITTA, Makoto AKABANE, Masato SHIMAKAWA, Nobuhide YAMAZAKI, Erika KOBAYASHI

William S. Frommer
Registration No. 25,506
FROMMER LAWRENCE & HAUG LLP
745 Fifth Avenue
New York, New York 10151
Tel. (212) 588-0800

SYNCHRONIZATION CONTROL APPARATUS AND METHOD, AND RECORDING MEDIUM

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to synchronization control apparatuses, synchronization control methods, and recording media. For example, the present invention relates to a synchronization control apparatus, a synchronization control method, and a recording medium suited to a case in which synthesized-voice outputs are synchronized with the operations of a portion which imitates the motions of an organ of articulation and which is provided for the head of a robot.

2. Description of the Related Art

Some robots which imitate human beings or animals have movable portions (such as a portion similar to a mouth which opens or closes when the jaws open and close) which imitate mouths, jaws, and the like. Others output voices while operating mouths, jaws, and the like.

When such robots operate the mouths and the like correspondingly to uttered words such that, for example, the mouths and the like have a shape in which human beings utter a sound of "a," at the output timing of a sound of "a," and have a shape in which human beings utter a sound of "i," at

the output timing of a sound of "i," the robots imitate human beings more real. However, such robots have not yet been created.

SUMMARY OF THE INVENTION

The present invention has been made in consideration of the foregoing condition. Accordingly, an object of the present invention is to implement a robot which imitates a human being more real in a way in which the operation of a portion which imitates an organ of articulation corresponds to uttered words generated by voice synthesis at utterance timing.

The foregoing object is achieved in one aspect of the present invention through the provision of a synchronization control apparatus for synchronizing the output of a voice signal and the operation of a movable portion, including phoneme-information generating means for generating phoneme information formed of a plurality of phonemes by using language information; calculation means for calculating a phoneme continuation duration according to the phoneme information generated by the phoneme-information generating means; computing means for computing the operation period of the movable portion according to the phoneme information generated by the phoneme-information generating means; adjusting means for adjusting the phoneme continuation

duration calculated by the calculation means and the operation period computed by the computing means; synthesized-voice-information generating means for generating synthesized-voice information according to the phoneme continuation duration adjusted by the adjusting means; synthesizing means for synthesizing the voice signal according to the synthesized-voice information generated by the synthesized-voice-information generating means; and operation control means for controlling the operation of the movable portion according to the operation period adjusted by the adjusting means.

The synchronization control apparatus may be configured such that the adjusting means compares the phoneme continuation duration and the operation period corresponding to each of the phonemes and performs adjustment by substituting whichever is the longer for the shorter.

The synchronization control apparatus may be configured such that the adjusting means performs adjustment by synchronizing at least one of the start timing and the end timing, of the phoneme continuation duration and the operation period corresponding to any of the phonemes.

The synchronization control apparatus may be configured such that the adjusting means performs adjustment by substituting one of the phoneme continuation duration and the operation period corresponding to all of the phonemes,

for the other.

The synchronization control apparatus may be configured such that the adjusting means performs adjustment by synchronizing at least one of the start timing and the end timing, of the phoneme continuation duration and the operation period corresponding to each of the phonemes, and by placing no-process periods at lacking intervals.

The synchronization control apparatus may be configured such that the adjusting means compares the phoneme continuation duration and the operation period corresponding to all of the phonemes and performs adjustment by extending whichever is the shorter in proportion.

The synchronization control apparatus may be configured such that the operation control means controls the operation of the movable portion which imitates the operation of an organ of articulation of an animal.

The synchronization control apparatus may further comprise detection means for detecting an external force operation applied to the movable portion.

The synchronization control apparatus may be configured such that at least one of the synthesizing means and the operation control means changes a process currently being executed, in response to a detection result obtained by the detection means.

The synchronization control apparatus may be a robot.

The foregoing object is achieved in another aspect of the present invention through the provision of a synchronization control method of synchronizing the output of a voice signal and the operation of a movable portion, including a phoneme-information generating step of generating phoneme information formed of a plurality of phonemes by using language information; a calculation step of calculating a phoneme continuation duration according to the phoneme information generated in the phoneme-information generating step; a computing step of computing the operation period of the movable portion according to the phoneme information generated in the phoneme-information generating step; an adjusting step for adjusting the phoneme continuation duration calculated in the calculation step and the operation period computed in the computing step; a synthesized-voice-information generating step of generating synthesized-voice information according to the phoneme continuation duration adjusted in the adjusting step; a synthesizing step of synthesizing the voice signal according to the synthesized-voice information generated in the synthesized-voice-information generating step; and an operation control step of controlling the operation of the movable portion according to the operation period adjusted in the adjusting step.

The foregoing object is achieved in still another

aspect of the present invention through the provision of a recording medium storing a computer-readable program for synchronizing the output of a voice signal and the operation of a movable portion, the program including a phoneme-information generating step of generating phoneme information formed of a plurality of phonemes by using language information; a calculation step of calculating a phoneme continuation duration according to the phoneme information generated in the phoneme-information generating step; a computing step of computing the operation period of the movable portion according to the phoneme information generated in the phoneme-information generating step; an adjusting step for adjusting the phoneme continuation duration calculated in the calculation step and the operation period computed in the computing step; a synthesized-voice-information generating step of generating synthesized-voice information according to the phoneme continuation duration adjusted in the adjusting step; a synthesizing step of synthesizing the voice signal according to the synthesized-voice information generated in the synthesized-voice-information generating step; and an operation control step of controlling the operation of the movable portion according to the operation period adjusted in the adjusting step.

In a synchronization control apparatus, a

synchronization control method, and a program stored in a recording medium according to the present invention, phoneme information formed of a plurality of phonemes is generated by using language information, and a phoneme continuation duration is calculated according to the generated phoneme information. The operation period of a movable portion is also computed according to the generated phoneme information. The calculated phoneme continuation duration and the computed operation period are adjusted, synthesized-voice information is generated according to the adjusted phoneme continuation duration, and a voice signal is synthesized according to the generated synthesized-voice information. In addition, the operation of the movable portion is controlled according to the adjusted operation period.

As described above, according to a synchronization control apparatus, a synchronization control method, and a program stored in a recording medium of the present invention, phoneme information formed of a plurality of phonemes is generated by using language information, a phoneme continuation duration and the operation period of a movable portion are calculated according to the generated phoneme information, the phoneme continuation duration and the operation period are adjusted, and the operation of the movable portion is controlled according to the adjusted operation period. Therefore, a word to be uttered by voice

synthesis at utterance timing can be synchronized with the operation of a portion which imitates an organ of articulation, and a more real robot is implemented.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram showing an example structure of a section controlling the operation of a portion which imitates an organ of articulation and controlling the voice outputs of a robot to which the present invention is applied.

Fig. 2 is a view showing example phoneme information and an example phoneme continuation duration.

Fig. 3 is a view showing example articulation-operation instructions and example articulation-operation periods.

Fig. 4 is a view showing an example of adjusted phoneme continuation durations.

Fig. 5 is a flowchart showing the operation of the robot to which the present invention is applied.

Figs. 6A and 6B show an example of a phoneme continuation duration and that of an articulation-operation period corresponding to each other, respectively.

Fig. 7 is a view showing the phoneme continuation duration and the articulation-operation period adjusted by a first method.

Fig. 8 is a view showing the phoneme continuation duration and the articulation-operation period adjusted by a

second method.

Figs. 9A and 9B show the phoneme continuation duration and the articulation-operation period adjusted by a third method, respectively.

Fig. 10 is a view showing the phoneme continuation duration and the articulation-operation period adjusted by a fourth method.

Fig. 11 is a view showing the phoneme continuation duration and the articulation-operation period adjusted by a fifth method.

Figs. 12A and 12B show examples in which phoneme information is synchronized with the operations of portions other than the organs of articulation.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Fig. 1 shows an example structure of a section controlling the operation of a portion which imitates an organ of articulation, such as jaws, lips, a throat, a tongue, or nostrils, and controlling the voice outputs of a robot to which the present invention is applied. This example structure is, for example, provided for the head of the robot.

An input section 1 includes a microphone and a voice recognition function (neither part shown), and converts a voice signal (words which the robot is made to repeat, such

as "konnichiwa" (meaning hello in Japanese), or words spoken to the robot) input to the microphone to text data by the voice recognition function and sends it to a voice-language-information generating section 2. Text data may be externally input to the voice-language-information generating section 2.

When the robot has a dialogue, the voice-language-information generating section 2 generates the voice language information (indicating a word to be uttered) of a word to be uttered as a response to the text data input from the input section 1, and outputs it to a control section 3. The voice-language-information generating section 2 outputs the text data input from the input section 1 as is to the control section 3 when the robot is made to perform repetition. Voice language information is expressed by text data, such as Japanese Kana letters, alphabetical letters, and phonetic symbols.

The control section 3 controls a drive 11 so as to read a control program stored in a magnetic disk 12, an optical disk 13, a magneto-optical disk 14, or a semiconductor memory 15, and controls each section according to the read control program.

More specifically, the control section 3 sends the text data input as the voice language information from the voice-language-information generating section 2, to a voice

synthesizing section 4; sends phoneme information output from the voice synthesizing section 4, to an articulation-operation generating section 5; and sends an articulation-operation period output from the articulation-operation generating section 5 and the phoneme information and a phoneme continuation duration output from the voice synthesizing section 4, to a voice-operation adjusting section 6. The control section 3 also sends an adjusted phoneme continuation duration output from the voice-operation adjusting section 6, to the voice synthesizing section 4, and an adjusted articulation-operation period output from the voice-operation adjusting section 6 to an articulation-operation executing section 7. The control section 3 further sends synthesized-voice data output from the voice synthesizing section 4, to a voice output section 9. The control section 3 furthermore halts, resumes, or stops the processing of the articulation-operation executing section 7 and the voice output section 9 according to detection information output from an external sensor 8.

The voice synthesizing section 4 generates phoneme information ("KOXNICHIWA" in this case) from the text data (such as "konnichiwa") output from the voice-language-information generating section 2 as voice language information, which is input from the control section 3, as shown in Fig. 2; calculates the phoneme continuation

duration of each phoneme; and outputs it to the control section 3. The voice synthesizing section 4 also generates synthesized voice data according to the adjusted phoneme continuation duration output from the voice-operation adjusting section 6, which is input from the control section 3. The generated synthesized voice data includes synthesized-voice data generated according to a rule, which is generally known, and data reproduced from recorded voices.

The articulation-operation generating section 5 calculates the articulation-operation instruction (instruction for instructing the operation of a portion which imitates each organ of articulation) corresponding to each phoneme and an articulation-operation period indicating the period of the operation, as shown in Fig. 3, according to the phoneme information output from the voice synthesizing section 4, which is input from the control section 3, and outputs them to the control section 3. In an example shown in Fig. 3, jaws, lips, a throat, a tongue, and nostrils serve as organs 16 of articulation. Articulation-operation instructions include those for the up or down movement of the jaws, the shape change and the open or close operation of the lips, the front or back, up or down, and left or right movements of the tongue, the amplitude and the up or down movement of the throat, and a change in shape of the nose. An articulation-operation instruction may be

independently sent to one of the organs 16 of articulation. Alternatively, articulation-operation instructions may be sent to a combination of a plurality of organs 16 of articulation.

The voice-operation adjusting section 6 adjusts the phoneme continuation duration output from the voice synthesizing section 4 and the articulation-operation period output from the articulation-operation generating section 5, which are input from the control section 3, according to a predetermined method (details thereof will be described later), and outputs to the control section 3. When the phoneme continuation duration shown in Fig. 2 and the articulation-operation period shown in Fig. 3 are adjusted according to a method in which whichever is the longer is substituted for the shorter for each phoneme in the phoneme continuation duration and the articulation-operation period, for example, the phoneme continuation duration of each of the phonemes "X," "I," and "W" is extended so as to be equal to the corresponding articulation-operation period.

The articulation-operation executing section 7 operates an organ 16 of articulation according to an articulation-operation instruction output from the articulation-operation generating section 5 and the adjusted articulation-operation period output from the articulation-operation adjusting section 6, which are input from the control section 3.

The external sensor 8 is provided, for example, inside the mouth, which is included in the organ 16 of articulation, detects an object inserted into the mouth, and outputs detection information to the control section 3.

The voice output section 9 makes a speaker 10 produce the voice corresponding to the synthesized voice data output from the voice synthesizing section 4, which is input from the control section 3.

The organ 16 of articulation is a movable portion provided for the head of the robot, which imitates jaws, lips, a throat, a tongue, nostrils, and the like.

The operation of the robot will be described next by referring to a flowchart shown in Fig. 5. In step S1, a voice signal input to the microphone of the input section 1 is converted to text data and sent to the voice-language-information generating section 2. In step S2, the voice-language-information generating section 2 outputs the voice language information corresponding to the text data input from the input section 1, to the control section 3. The control section 3 sends the text data (for example, "konnichiwa") serving as the voice language information input from the voice-language-information generating section 2, to the voice synthesizing section 4.

In step S3, the voice synthesizing section 4 generates phoneme information (in this case, "KOXNICHIWA") from the

text data serving as the voice language information output from the voice-language-information generating section 2, which is sent from the control section 3; calculates the phoneme continuation duration of each phoneme; and outputs to the control section 3. The control section 3 sends the phoneme information output from the voice synthesizing section 4, to the articulation-operation generating section 5.

In step S4, the articulation-operation generating section 5 calculates the articulation-operation instruction and articulation-operation period corresponding to each phoneme according to the phoneme information output from the voice synthesizing section 4, which is sent from the control section 3, and outputs them to the control section 3. The control section 3 sends the articulation-operation period output from the articulation-operation generating section 5 and the phoneme information and the phoneme continuation duration output from the voice synthesizing section 4, to the voice-operation adjusting section 6.

In step S5, the voice-operation adjusting section 6 adjusts the phoneme continuation duration output from the voice synthesizing section 4 and the articulation-operation period output from the articulation-operation generating section 5, which are sent from the control section 3, according to a predetermined rule, and outputs to the

control section 3.

First to fifth methods for adjusting the phoneme continuation duration and the articulation-operation period will be described here by referring to Figs. 6A, 6B, 7, 8, 9A, 9B, 10, and 11. In the following description, it is assumed that the phoneme continuation duration generated in step S3 is shown in Fig. 6A and the articulation-operation period generated in step S4 is shown in Fig. 6B.

In the first method, the phoneme continuation duration and the articulation-operation period of each phoneme are compared, and whichever is the longer is used to substitute for the shorter. Fig. 7 shows an adjustment result obtained by the first method. In examples shown in Figs. 6A and 6B, since the phoneme continuation duration of each of the phonemes "K," "CH," and "W" is longer than the corresponding articulation-operation period, the articulation-operation period is substituted for the phoneme continuation duration as shown in (B) of Fig. 7. Conversely, since the articulation-operation period of each of the phonemes "O," "X," "N," "I," "I," and "A" is longer than the corresponding phoneme continuation duration, the phoneme continuation duration is substituted for the articulation-operation period as shown in (A) of Fig. 7.

In the second method, the start timing or the end timing of any phoneme is synchronized. Fig. 8 shows an

adjustment result obtained by the second method. When synchronization is achieved at the start timing of the phoneme "X," as shown in Fig. 8, data lacks before the starting timing of the phoneme continuation duration of the phoneme "K" and after the end timing of the phoneme continuation duration of the phoneme "A." Adjustment is achieved such that voices are not uttered at the data-lacked portions and only articulation operations are performed. The user may specify the phoneme at which the start timing is synchronized. Alternatively, the control section 3 may determine according to a predetermined rule.

In the third method, either the phoneme continuation duration or the articulation-operation period is used for all phonemes. Fig. 9 shows an adjustment result obtained by the third method in a case in which the articulation-operation period has priority and the articulation-operation period is substituted for the phoneme continuation duration for all phonemes. The user may specify which of the phoneme continuation duration and the articulation-operation period has priority. Alternatively, the control section 3 may select either of them according to a predetermined rule.

In the fourth method, the start timing or the end timing of each phoneme is synchronized between the phoneme continuation duration and the articulation-operation period, and blanks are placed at lacking periods of time (indicating

periods when neither utterance nor an articulation operation is performed). Fig. 10 shows an adjustment result obtained by the fourth method. A blank is placed at a lacking period of time generated before the start timing of the phoneme "K" in the articulation-operation period as shown in (B) of Fig. 10, and blanks are placed at lacking periods of time generated before the starting timing of the phonemes "O," "X," "N," and "I" in the phoneme continuation duration, as shown in (A) of Fig. 10.

In the fifth method, the start timing or the end timing of the phoneme located at the center of the phoneme information is synchronized, the entire phoneme continuation duration and the entire articulation-operation period are compared, and the shorter period is extended so that it has the same length as the longer. More specifically, for example, as shown in Fig. 11, the start timing of the phoneme "I" located at the center of the phoneme information "KOXNICHIWA" is synchronized and the phoneme continuation duration is extended to 550 ms since the entire phoneme continuation duration (300 ms) is shorter in time than the articulation-operation period (550 ms). Further specifically, the phoneme continuation duration of each of the phonemes "K," "O," "X," and "N," which are located before the phoneme "I," is twice (= 300/150) extended, and the phoneme continuation duration of each of the phonemes

"I," "CH," "I," "W," and "A," which are located after the phoneme "I," is extended by a factor of 1.25 (= 250/200).

As described above, the phoneme continuation duration and the articulation-operation period are adjusted by one of the first to fifth methods, or by a combination of the first to fifth methods, and sent to the control section 3.

Back to Fig. 5, in step S6, the control section 3 sends the adjusted phoneme continuation duration output from the voice-operation adjusting section 6, to the voice synthesizing section 4, and sends the adjusted articulation-operation period output from the voice-operation adjusting section 6 and the articulation-operation instruction output from the articulation-operation generating section 5, to the articulation-operation executing section 7. The voice synthesizing section 4 generates synthesized voice data according to the adjusted phoneme continuation duration output from the voice-operation adjusting section 6, which is input from the control section 3, and outputs it to the control section 3. The control section 3 also sends the synthesized voice data output from the voice synthesizing section 4 to the voice output section 9. The voice output section 9 makes the speaker produce the voice corresponding to the synthesized voice data output from the voice synthesizing section 4, which is input from the control section 3. In synchronization with this operation, the

articulation-operation executing section 7 operates the organ 16 of articulation according to the articulation-operation instruction output from the articulation-operation generating section 5 and the adjusted articulation-operation period output from the voice-operation adjusting section 6, which are input from the control section 3.

Since the robot is operated as described above, the robot imitates the utterance operations of human beings and animals more natural.

When the external sensor 8 detects an object inserted into the mouth, which is included in the organ 16 of articulation, during the process of step S6, detection information is sent to the control section 3. The control section 3 halts, resumes, or stops the processing of the articulation-operation executing section 7 and the voice output section 9 according to the detection information. With this operation, since a voice cannot be uttered when the object is inserted into the mouth, reality is enhanced. In addition to a case in which the detection information is sent from the external sensor 8, when the operation of the organ 16 of articulation is disturbed by some external force, the processing of the voice output section 9 may be halted, resumed, or stopped.

In such a control, utterance processing is changed in response to a change of an articulation operation.

Conversely, control may be executed such that an articulation operation is changed in response to a change of utterance processing, such as in a case in which an articulation operation is immediately changed when a word to be uttered is suddenly changed.

In the present embodiment, the output of the voice-language-information generating section 2 is set to text data, such as "konnichiwa." It may be phoneme information, such as "KOXNICHIWA."

The present invention can also be applied to a case in which the phonemes of an uttered word are synchronized with the operation of a portion other than the organs of articulation. In other words, the present invention can be applied, for example, to a case in which the phonemes of an uttered word are synchronized with the operation of a neck or the operation of a hand, as shown in Fig. 12.

In addition to robots, the present invention can further be applied to a case in which the phonemes of words uttered by a character expressed by computer graphics are synchronized with the operation of the character.

The above-described series of processing can be executed by software as well as by hardware. When the series of processing is executed by software, the program constituting the software is installed from a recording medium into a computer built in a special hardware or into a

general-purpose personal computer which executes various functions with installed various programs.

This recording medium can be a package medium storing the program and distributed to the user to provide the program separately from the computer, such as a magnetic disk 12 (including a floppy disk), an optical disk 13 (including a compact disk-read only memory (CD-ROM) and a digital versatile disk (DVD)), an magneto-optical disk 14 (including a Mini disk (MD)), or a semiconductor memory 15. In addition, the recording medium can be a ROM or a hard disk storing the program and distributed to the user in a condition in which it is placed in the computer in advance.

In the present specification, steps describing the program which is stored in a recording medium include processes executed in a time-sequential manner according to the order of descriptions and also include processes executed not necessarily in a time-sequential manner but executed in parallel or independently.